AD NUMBER
ADA801494
CLASSIFICATION CHANGES
TO: **UNCLASSIFIED**
FROM: **RESTRICTED**
LIMITATION CHANGES
TO: **Approved for public release; distribution is unlimited.**
FROM: **Distribution authorized to U.S. Gov't. agencies and their contractors; Administrative/Operational Use; 05 JUL 1944. Other requests shall be referred to Office of Scientific Research and Development, Washington, DC 20301.**
AUTHORITY
OSRD list no. 9 dtd 18-21 Feb 1946; OTS index dtd Jun 1947

The U.S. GOVERNMENT

IS ABSOLVED

FROM ANY LITIGATION WHICH MAY ENSUE FROM THE CONTRACTORS IN-FRINGING ON THE FOREIGN PATENT RIGHTS WHICH MAY BE INVOLVED.

RESTRICTED

REPORT ON

TRAINING STUDIES IN VOICE COMMUNICATION:

I. CAN INTELLIGIBILITY OF VOICE COMMUNICATION BE INCREASED BY TRAINING IN VOICE TECHNIQUE?

OSRD Report No. 3862

July 5, 1944

Applied Psychology Panel, NDRC

RESTRICTED

RESTRICTED

Applied Psychology Panel, NDRC
Project N-109(1), SC-67: VOICE COMMUNICATION

Report No. 8, July 5, 1944

REPORT ON TRAINING STUDIES IN VOICE COMMUNICATION: I. CAN INTELLIGIBILITY OF VOICE COMMUNICATION BE INCREASED BY TRAINING IN VOICE TECHNIQUE?

Prepared by

James F. Curtis

Assisting Laboratory Staff
I. P. Brackett
Henry M. Moser

Statistical Analysis
C. Hess Haagen

Approved for Distribution

W. S. Hunter

By
Chief, Applied Psychology Panel

Office of Scientific Research and Development
Contract OEMsr-830

OSRD Report No. 3862

Copy No. 176 of 200 copies

THE PSYCHOLOGICAL CORPORATION
New York, New York
George K. Bennett, Contractor's Technical Representative
John W. Black, Project Director

RESTRICTED

RESTRICTED

DISTRIBUTION LIST

Applied Psychology Panel

Dr. W. S. Hunter
Dr. G. K. Bennett
Dr. Leonard Carmichael
Dr. C. H. Graham
Mr. J. M. Stalnaker
Dr. M. S. Viteles
Dr. J. L. Kennedy, 5 copies
Dr. D. L. Wolfle, 5 copies

Consultant to the Panel

Dr. D. W. Bronk

Contractors' Technical Representatives and Project Directors, Panel Projects

Dr. J. G. Beebe-Center
Dr. W. C. Biel
Dr. W. J. Brogden
Dr. R. K. Campbell
Dr. J. H. Gorsuch
Dr. H. O. Gulliksen
Dr. W. E. Kappauf
Dr. F. S. Keller
Dr. T. L. Kelley
Dr. H. S. Langfeld
Dr. D. B. Lindsley
Dr. L. C. Mead
Dr. J. P. Nafe
Dr. H. W. Nissen
Dr. W. C. Shipley
Dr. K. R. Smith
Mr. H. A. Voss
Dr. C. H. Wedell

Army

Adjutant General's Office
**Dr. W. V. Bingham
*Dr. Herman Copeland
**Col. G. R. Evans
**Lt. Col. M. W. Richardson

Assistant Chief of Staff, G-3
Attn: Chief Training Group

Headquarters, Army Air Forces
Assistant Chief of Air Staff, Tr'ng.
AC/AS, Training
Attn: Communications Branch
*Lt. Col. J.C. Courtney, Jr.

AC/AS, Training
Attn: Radar Coordination Office
*Col. W. H. Tetley
Office of the Air Communications Officer
Attn: Radio Navigation Branch
Capt. L. A. McClelland
Office of the Air Communications Officer
Attn: Training Liaison
Col. Clay Hoppough
Assistant Chief of Air Staff, Operations, Commitments and Requirements
AC/AS, O. C. & R.
Attn: Bombardment Branch
Maj. R. A. Trenkle
Management Control
Attn: Operations Analysis
Lt. Col. W. B. Leach
The Air Surgeon
Attn: Research Division
Col. L. E. Griffis
The Air Surgeon
Attn: **Col. J. C. Flanagan

Office of the Secretary of War
Attn: Mr. M. E. Strieby

Office of the Chief Signal Officer
Plans and Operations Division (5 copies)
Attn: Operational Research Branch
**Dr. Don Lewis
Chief, Personnel and Training Service
Chief, Engineering and Technical Service
Attn: Aircraft Radio Branch
Commanding General, Eastern Signal Corps School, Fort Monmouth, N. J.

Headquarters, Army Ground Forces
Assistant Chief of Staff, G-3
Signal Section
**Capt. Dwight L. Emmel

Communications Control Division, Headquarters, Bolling Field, D. C.
Air Communications Office
Attn: Maj. Virgil Clapp

Army Air Forces Tactical Center, Orlando, Florida
Commanding General
Col. O. L. Grover

RESTRICTED

DISTRIBUTION LIST (Cont'd)

Army Air Forces Training Command,
Fort Worth, Texas
Commanding General
Assistant Chief of Staff, A-3
Attn: Pilot Training
Col. P. E. Todd
Assistant Chief of Staff, A-3
Attn: Air Crew Non-Pilot Training
Col. H. Crutcher
Assistant Chief of Staff, A-3
Attn: Technical Training
Col. C. H. Rees
Assistant Chief of Staff, A-3
Attn: Training Aids Section
Maj. C. M. Peterson
Assistant Chief of Staff, A-3
Attn: Signal Section
Col. H. L. Bennett

Army Air Forces Central Flying Training
Command, Randolph Field, Texas
Commanding General
Assistant Chief of Staff, A-3
Attn: Ground School Section
Col. J. H. Montfort

Army Air Forces Central Instructors
School, Randolph Field, Texas
Office of the Director
Ground School
Attn: Maj. C. H. Rhoadman
Technical Advisory Group
Attn: Maj. Thomas A. Rousse
Psychological Research Unit
Attn: Capt. Neil Miller

School of Aviation Medicine,
Randolph Field, Texas
Research Division
Attn: Lt. Col. Paul Campbell
Research Division
Attn: Capt. Ben H. Senturia

Waco Army Air Field, Texas
Attn: Col. Bob Arnold

Army Air Forces Training Aids
Division, New York, N. Y.
Commanding Officer

Army Air Forces Instructors School -
Instruments, Bryan, Texas
Commanding Officer

Army Air Forces Eastern Flying Training
Command, Maxwell Field, Alabama
Commanding General
Assistant Chief of Staff, A-3

Army Air Forces Western Flying Training
Command, Santa Ana Field, California
Commanding General

1st Air Force, Mitchell Field, New York
Commanding General

2nd Air Force, Colorado Springs, Colorado
Commanding General

3rd Air Force, Drew Field, Florida
Commanding General

4th Air Force, San Francisco, California
Commanding General

Army Air Forces Central Instructors
School - Gunnery, Buckingham Army Air
Field, Fort Myers, Florida
Director of Training

Army Air Forces Central Instructors
School - Flexible Gunnery, Laredo Army
Air Field, Laredo, Texas
Research Division
Maj. Nicholas Hobbs

Army Air Forces Central Instructors
School - Navigation, Selman Army Air
Field, Monroe, Louisiana
Commanding Officer
Director of Training
Psychological Research Unit
Attn: Capt. L. F. Carter

Army Air Forces Central Instructors
School - Bombardiers, Midland Army Air
Field, Midland, Texas
Director of Training
Psychological Research Unit
Attn: Maj. Edward Kemp

DISTRIBUTION LIST (Cont'd)

Army Air Forces Central Instructors School - B-17 Pilots, Lockbourne Army Air Field, Ohio
Director of Training

Army Air Forces Central Instructors School - B-24 Pilots, Smyrna Army Air Field, Tennessee
Director of Training

Scott Field, Illinois
Director of Training - Technical
Attn: Lt. W. E. Biggs

President, Army Air Forces Board
Orlando, Florida
Attn: Communications Division

Coast Guard

Training Division
Lt. J. E. McVey

Marine Corps

Division of Aviation
**Capt. Paul Martin

Personnel Department
**Maj. R. W. Rickert

Marine Liaison Officer,
Combat Operation Dept. Orlando, Fla.
Capt. Paul W. Smith

Navy

Office of the Commander-in-Chief
Readiness Division
**Comdr. D. C. Beard
Lt. S. H. Britt
**Capt. J. H. Thach, Jr.

VCNO (Air)
*Comdr. John W. Bowan, Jr.

Interior Control Board
*Lt. G. W. Dyson

Bureau of Medicine and Surgery
Medical Research Division
**Comdr. J. G. Jenkins

Bureau of Naval Personnel
Enlisted Distribution Division
Lt. Comdr. John Blann
Lt. Newell C. Kephart, 6 copies
Planning and Control Division
**Lt. Comdr. J. C. O'Brien
Quality Control Division
**Capt. W. E. Hennigar
Standards and Curriculum Division
**Comdr. C. R. Adams

Bureau of Ships
*Lt. Comdr. C. G. Grimes

Ordnance and Gunnery School
Lt. Comdr. A. L. Shepherd

Amphibious Training Command, Atlantic
**Capt. P. E. McDowell

U. S. Submarine Base, New London
Medical Research Laboratory
**Capt. C. W. Shilling

NavTraSta, Bainbridge, Maryland
Lt. David Braswell

NavTraSta, Fort Lauderdale, Florida
Commanding Officer

NavAirSta, Corpus Christi, Texas
Aviation Safety Board
Lt. (jg) F. H. Sanford

NavAirSta, New Orleans, Louisiana
Naval Flight Instructors School
Lt. (jg) W. H. Wilke

NavAirSta, Atlanta, Georgia
Instrument Flight Instructors School
Lt. E. T. Curry

Nav Aviation TraSta, Pensacola, Florida
Aviation Psychology
Lt. Max Steer

Operational Training Command
Atlantic Commanding Officer

National Research Council

Dr. L. H. Weed

DISTRIBUTION LIST (Cont'd)

National Defense Research Committee

Executive Secretary, 24 copies

Division 6
Dr. Adelbert Ford
Dr. G. P. Harnwell
Dr. W. D. Neff

Section 7.4
Dr. S. W. Fersberger

Section 17.3
Dr. S. S. Stevens

OSRD Liaison Office for transmittal to

R. A. F. Delegation
Wing Comdr. P. A. Lee

Royal Navy
Surgeon Capt. R. W. Mussen

R. C. A. F., Toronto, Ontario
Squadron Leader A. E. Jarvis

**General Liaison Officers
*Project Liaison Officers

DISTRIBUTION LIST (Cont'd)

National Defense Research Committee

Executive Secretary, 24 copies

Division 6
Dr. Adelbert Ford
Dr. G. P. Harnwell
Dr. W. D. Neff

Section 7.4
Dr. S. W. Feriberger

Section 17.3
Dr. S. S. Stevens

OSRD Liaison Office for transmittal to

R. A. F. Delegation
Wing Comdr. P. A. Lee

Royal Navy
Surgeon Capt. R. W. Mussen

R. C. A. F., Toronto, Ontario
Squadron Leader A. E. Jarvis

**General Liaison Officers
*Project Liaison Officers

SUMMARY

The Voice Communication Laboratory was established to investigate the possibility of improving communication efficiency in Army Air Forces, through training personnel in proper use of equipment, use of standard radiotelephone procedure and voice technique, and to develop procedures and devices for implementing such training. A number of experiments have been conducted investigating various phases of this problem. This report presents data from three selected experiments to show the results obtained from training student pilot subjects in voice technique.

In all of the training experiments of the Voice Communication Laboratory, the criterion for evaluating results has been a word intelligibility test. This testing technique has been patterned after the word "articulation" tests which have been used extensively for testing efficiency of communication equipment.

The following standard communication equipment in use in Army Air Forces was employed: T-17 (hand-held) microphones; BC-347-C interphone amplifiers; HS-23 (R-14 earphones) or HS-33 (ANB-H-1 earphones) headsets. All tests were run in a simulated airplane noise of approximately 110 db average level.

The design of the typical training experiment is that of an initial test, then a period of training, followed by a re-test. The index of improvement of an experimental group is its mean gain in test score, from the initial to the final test, adjusted for differences between groups in initial test score. This adjustment is necessary because of the relationship existing between initial level and improvement, and because groups of subjects are not matched in initial ability. Since this relationship is rectilinear, analysis of covariance provides a convenient means of making the appropriate adjustment.

Control groups are frequently used in the training experiments. These groups are tested at the same times as the trained groups, but receive no training in the laboratory between the initial and final tests.

Results:

1. Training for a period as short as two hours produces gains in intelligibility score which are greater than that for a control group by amounts which are statistically significant at high levels of confidence.

2. Further training up to six hours shows additional improvement. No experiment involving a training period of more than six hours is reported.

3. Significant differences exist between training procedures, indicating a need for careful working out of most advantageous procedures.

4. A breakdown of the data into quartiles on the basis of the initial test scores shows that the subjects who scored lowest on the initial tests make the most marked improvement. A considerable spread in mean score for the four quartiles is found on the initial test. By the time of the final test, however, the poorer subjects have improved sufficiently that the differences between the quartiles are much smaller.

Attention is called to the fact that the student pilot subjects used in these experiments already possessed relatively good speech habits prior to training. Subjective judgment indicates that they are well above the average of the general population in this respect. The results shown by training were not, therefore, the result of correcting generally bad and faulty articulation and pronunciation habits, or poor voice usage, but of training the subjects to use their voices to good advantage in a relatively unique speaking situation, which involves certain difficulties and requires some special skill and technique.

INTRODUCTION

Various surveys of communication conditions in Army Air Forces have revealed that voice communication is, at times, far from satisfactory. This has been found despite the fact that a great deal of effort has been directed at improving communication equipment, and that improvements have been made in earphones, earphone cushions, microphones, interphone amplifiers, and radio transmitting and receiving apparatus. The importance of good communication to pilots and air crew members may be indicated by the fact that, when experienced RAF pilots were asked to rate factors which helped to relieve fatigue during heavy bombardment missions, they ranked good intercommunication as one of the first three. The same pilots commented that nothing was more annoying than to be requested to repeat messages, or to be obliged to ask for repeats of messages.

The fact is that, until recently, scant heed was paid to one of the very important factors in communication. Voice communication is, after all, a chain which may be represented schematically as follows:

The Speaker ----> The Equipment ----> The Listener

Although the equipment link in this chain has received the lion's share of attention, the listener and speaker are scarcely less important. No ingenious improvement in equipment can produce its maximum effect unless the speaker and listener also operate with maximum effectiveness.

The Voice Communication Laboratory was established under a relatively broad directive to investigate and develop procedures, devices, and materials for training Army Air Forces personnel in the use of voice communication equipment. Three specific types of skills are mentioned as those which should be unified under such a training program, viz., (a) speaking and listening in the presence of noise, interference, distortion, etc.; (b) use of standard radiotelephone procedures; (c) manipulation of equipment, e. g., tuning, microphone and earphone technique, etc.

The major portion of the efforts of the Voice Communication Laboratory, to date, have been directed at the first of these, training in speaking and listening in the presence of noise. There have been several reasons for this particular direction of effort:

(1) This was the skill about which least was known. It was well known that individuals differ markedly in their skill in speaking intelligibly under conditions of noise and interference, and in their ability to hear speech accurately under such conditions, but little was known concerning the factors which produce such differences.

(2) Training procedures in this type of skill had not been developed to any extent. Efforts had been made in this direction by interested individuals in scattered locations,

but no systematic program had been undertaken and no evaluation of results of such training were available. Training methods and procedures with respect to the other skills were thought to be more straightforward and, hence, less in need of experimental investigation and trial. Teaching of standard procedure and its use, for example, could be done with procedures and methods quite similar to those used in other types of training.

(5) There was some question as to whether effective training in speaking and listening could be provided within the limits of time required by the already crowded curricula of AAF training schools. It seemed important to obtain a definite answer one way or the other as early as possible.

METHOD OF EVALUATING TRAINING

General

The method of evaluating results which has been used was specifically chosen to measure the effectiveness of the type of training referred to above, i.e., training to develop skill in speaking and listening under conditions of noise and interference similar to those experienced in military aircraft. The criterion measurement for all experiments has been a word intelligibility test.

The basic methodology of such tests is not new. Under the name of "articulation tests," they have been used extensively for many years by researchers of the Bell Telephone Laboratories for measuring the relative efficiency of telephone and other electrical communication circuits, under various conditions.* In recent years, the Harvard Psycho-Acoustic Laboratory has also made extensive use of such tests to measure the relative efficiency of various kinds of voice communication equipment for the Armed Forces,** and such tests have been used for testing equipment by other groups, including Aircraft Radio Laboratory, Wright Field. The use of such tests as a criterion for evaluating the efficiency of communication equipment is well established and accepted. In the work of the Voice Communication Laboratory, these tests have simply been adapted to use as a criterion for evaluating the performance of groups of speakers and listeners. The term "intelligibility test" has been used instead of "articulation test" in this laboratory as being more descriptive of the function under study, since it is the improvement in audibility or intelligibility of communication that is the matter of chief interest.

---

* Cf. Fletcher, H., Speech and Hearing, D. Van Nostrand Company, New York, 1929, Part Four, Chaps. III-VII. Also
Fletcher, H. and Steinberg, J. C., Articulation Testing Methods, Bell System Tech. Jour., VIII, 1929, 806-854.

** Cf. Articulation Testing Methods, National Research Council, Committee on Sound Control, Feb. 1, 1942.

Description of Experimental Room and Testing Procedure

Figure 1 is a schematic diagram of the experimental room. It is equipped with an interphone system, made up of standard Signal Corps components used in Army aircraft. Twenty stations, each equipped with a headset and microphone, are located around three sides of the room. Two stations are located at the control table for the monitors. The interphone circuit is so arranged that all stations may be included on one circuit, or the stations may be divided among either two or three circuits. For all experiments reported here the basic components of the interphone system have been: T-17 (hand-held) microphones; BC-347-C interphone amplifiers; and either HS-23 (R-14 earphones) or HS-33 (ANB-H-1 earphones) headsets. In addition to the interphone system, an essential part of the room is the noise producing apparatus. It consists of an electronic noise generator and power amplifier, both located on the control table, and a loudspeaker for projecting the sound into the room. The noise is similar in spectrum to that found in typical military aircraft in flight and has an average level in the room of 110 db. Recording equipment, and any other devices used in training, may also be located on the control table.

As used in the typical training experiment, a test consists of twenty-four words read by a cadet speaker to a circuit of eight to ten cadet listeners. The listeners are provided with sheets of answer blanks on which they write the words as they hear them. The noise is turned on throughout the time that a test is being read. A speaker's intelligibility score consists of the percentage of correct responses written by the listeners for the twenty-four words read by the speaker. For example, if ten listeners write an average of eleven correct responses for a certain speaker, there would be a total of 110 correct responses out of a possible total of 240. The speaker's intelligibility score would be 110/240 or 46%.

The usual testing procedure is to divide the stations into two circuits and test two speakers simultaneously, one on each circuit. All subjects act as listeners except at the time when they are reading their own test lists. Since there is some variation in noise level from station to station, one station in each circuit is designated as the speaker's station. Each subject goes to the speaker's station on his circuit when his turn comes to read his test list. The noise level is thus a constant for all speakers within each circuit. The group of subjects tested on a single circuit are kept as nearly intact as possible from one test to another. Subjects are never changed back and forth from one circuit to another when scores from two sets of tests are to be compared. If, due to absence or elimination of cadets, subjects are present on one test, but not on another, scores are computed using only the papers of those listeners who were present for both of a speaker's tests. Thus, scores which are to be compared are always obtained with a constant personnel composing the listening group.

This testing arrangement and the use of short test lists (twenty-four words) have made it possible to test twenty to twenty-four subjects in one hour. This was necessary, if the testing procedure was not to occupy a disproportionately large amount of the available time of the subjects.

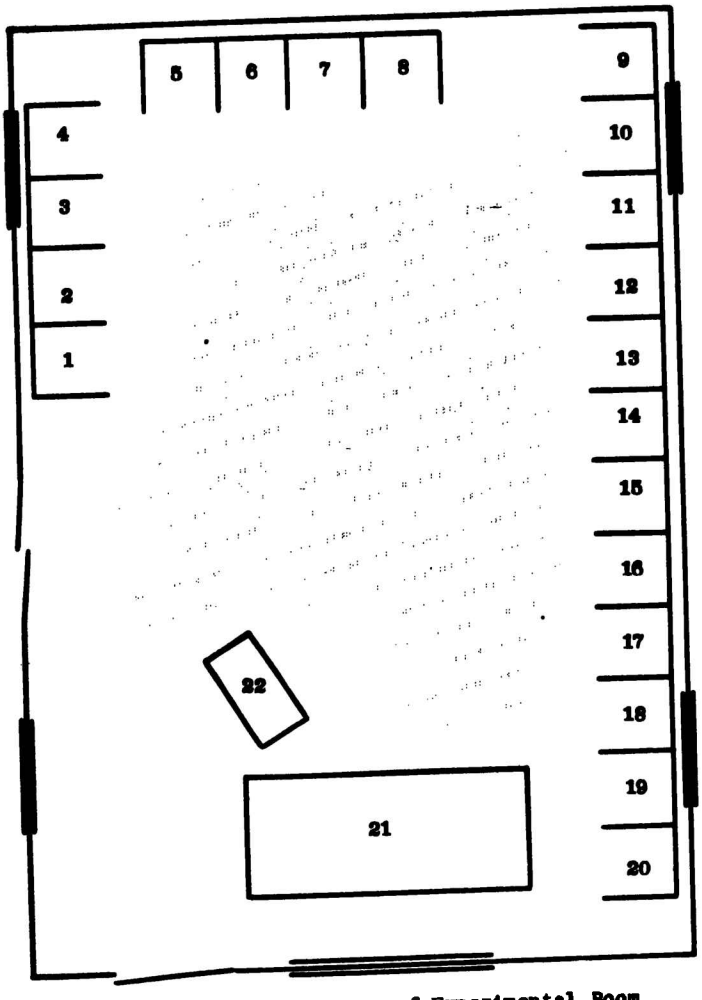

Figure 1. Schematic Diagram of Experimental Room

1-20 Student Stations
21 Control Table
22 Speaker for Producing Noise

The test lists were made up with considerable care, and analysis of results has demonstrated that, with groups of the size used in the training experiments, tests of this length yield mean scores of satisfactory reliability. A complete discussion of the construction of the test lists and of the statistics concerning their reliability is contained in a separate report.*

Basis for Choice of a Word Test

The choice of a word test, rather than a sentence test, or some other type, was based on several considerations, some of them of a very practical nature. A test was required that could be administered to sizeable groups within a short period, not more than one hour. The test had to be one that could be easily scored by workers without special skill. It had to be sufficiently reliable, and of sufficient discriminating power to provide reliable measurement of differences between groups and between tests and re-tests for a single group. Lastly, but by no means least important, a test was required that could be assumed to be reasonably valid. No test that completely fitted all of these requirements was available, and time did not permit spending months in developing one. A word test, while not perfect, seemed to satisfy these criteria reasonably well and was more readily adaptable to the particular needs of the experimental program than any other.

A sentence test, or other connected speech test, involves certain difficulties which make its use difficult, if not impossible, under the circumstances existing in the laboratory. In the first place, the administration time is excessive. Much more time is needed to test each speaker since it takes more time to speak each item. Much more time for test development would have been required. Many hundreds of test items would have had to be written, tried out, and analyzed for discriminating power before an adequate number of tests of equal difficulty would have been available. The scoring of most types of sentence tests is less convenient, and, in the hands of unskilled help, probably less reliable. All of these considerations make a sentence type test impractical for use in evaluating most training experiments.

Validity of a Word Test

A point sometimes questioned concerning a word test used for this purpose is its validity. Although there is little objective evidence available on the point, there is considerable logical evidence to support the validity of a word test as used in the experiments. The men speak and listen with standard equipment of the sort they will be called on to use when flying. They speak and listen under conditions of noise similar to those existing in military airplanes. A not inconsiderable part of practical voice communication in Air Forces is made up of single words, especially where messages are coded into code words or numbers. Even where connected speech is used, short phrases are the rule, and one or two words may carry the main content of the phrase. In other words, the test is a reasonably realistic job sample of the communication work which the men are actually called on to perform.

---

* Intelligibility Measurement--Techniques and Procedures Used by the Voice Communication Laboratory, in process of publication.

In addition to such logical grounds for assuming validity, there is a limited amount of objective data. The relationship between intelligibility scores obtained on sentence tests and scores obtained on word tests under similar conditions has been determined.* Data also is available with respect to the relationship between scores obtained for sentence tests and those for syllable tests.** In both cases, the relationships, though curvilinear, are positive and appear to be high. Although these data are based on differences in signal level, distortion, noise levels, etc., rather than individual differences among speakers, they do show that there is a positive relation between intelligibility as measured by connected speech tests, such as sentences, and as measured by tests which use more fractional elements of speech, such as words and syllables.

Probably the most serious objection to the use of a word test in evaluating the results of training in voice technique for communication work is that some elements, which may contribute to the relative intelligibility of connected speech, are not measured in a word test. This may well be true. However, in so far as it is applicable, it would seem to mean that the tests do not show the whole picture, rather than that the picture obtained from such tests is inaccurate. Where improvement, following training, is found through use of a word test, it may be that the test fails to measure all of the factors which have been affected by training, and that a more adequate sampling of these factors would indicate even more substantial improvement. On the other hand, it hardly seems likely that improvement obtained by testing with words would be shown to be spurious if a test providing a more complete sampling of speech factors were employed.

GENERAL METHODOLOGY OF TRAINING EXPERIMENTS

The general design of most of the training experiments has been that of an initial test, then a period of training, followed by a re-test. The index of improvement of a group is the gain in score between the initial and final tests. Ordinarily the subjects have been given the initial test during the second hour that they report to the laboratory. The first hour is usually devoted to satisfying the curiousity of the cadets as to why they are required to come to the laboratory, explaining in a very general way what will be expected of them, and giving preliminary instruction in the use of the equipment which will be employed in the testing. It has been the unanimous opinion of the laboratory that the time spent in such preliminary indoctrination was amply justified by better rapport with the subjects, and more adequate motivation for the testing and training work which followed.

The content of the training work, in the experiments to be discussed, has been directed almost entirely toward improvement in voice technique.

---

* Articulation Testing Methods, National Research Council Committee on Sound Control, February 1, 1942, P. 35.

** Fletcher, H. and Steinberg, J. C., Articulation Testing Methods, Bell System Technical Journal, VIII, 1929, P.45.

Subjects have been instructed in such matters as the loudness level of voice for greatest intelligibility, adequate rate of speech and phrasing of messages, good clear articulation and pronunciation, proper pitching of the voice for communication in noise, etc. Lecture and discussion have typically been held to the minimum necessary to give the subjects the information upon which to base intelligent practice. The greater portion of the class time was devoted to practice and drill, working toward developing these techniques of voice. For such practice, messages, phrased in standard radiotelephone procedure language, such as pilots and air crews are called on to speak and be able to understand, were used as much as possible. Use has been made of demonstration recordings to illustrate various faults and to show the difference between good and poor intelligibility of communication. Recording equipment has at times been used so that subjects could hear themselves, and arrive at some better estimate of how they sound to others when talking over the interphone or radio in noise. Throughout the practice and drill the instructor has acted as guide and critic to point out faults, make suggestions for improvement, and keep the practice moving as rapidly as possible. The instructors for the training experiments were men who had several years of college or university experience in teaching speech. However, except for the experimental program, in which procedures and training methods are being developed, it is not considered that instructors with such specialized experience and training will be needed. The laboratory now has, in process of publication, manuals, syllabi for instructors, etc., which should make it possible for any good teacher to do a successful job of giving this training. Experience in teaching speech might be helpful, but by no means prerequisite, to instructing in this training work.

Control groups have frequently been used in the experiments. These were groups which were tested initially at the same time as the groups to be given the experimental training treatments, were then dismissed, and were not seen again until the time of the final test some days or weeks later. As will be shown (see Figures 2 and 4), it has been usual to find that the control groups made improvement. This was to be expected, since they had, by the time of the second test, received training from at least three sources. One, they fly almost daily and hence make use of interphone and radio equipment in noise. Two, the ground school at the Waco Army Air Field devotes a portion of its course in Radio Communication to some training in voice procedure and technique. Three, the test situation itself provides a rather considerable learning experience. In the process of taking the test, the subjects spend an hour listening and speaking in noise. They hear nine or ten different cadets speak words over an interphone system in noise, and they no doubt form opinions as to why some are easier to understand than others. In addition, it is not improbable that cadets from the control groups pick up information from the cadets in the groups undergoing training. In a sense, then, such groups are not strictly control groups. They do, however, provide some base against which to evaluate the improvement of the experimentally trained groups, and any error made by comparing the controls and the trained groups is probably on the conservative side so far as evaluating the results of training is concerned.

One additional point concerning evaluation of results, not covered in the preceding section, ought, probably, to be discussed here. The system of testing used, wherein the subjects for training act as both speakers and listeners during the test, yields a score which is not only a function of the intelligibility of the words as uttered by the speakers, but also a function of the skill of the listeners in correctly identifying them and writing them down. In so far as the same persons act as listeners for both the initial and final tests; and in so far as their skill in listening is a constant from test to test, this provides no problem. The latter, however, is probably not true. It is not improbable that the subjects improve as listeners to some degree, as well as improving in voice technique. The improvement on the speaking end and the improvement on the listening end are confounded in the improvement scores actually obtained. The only way, whereby a pure measure of improvement in voice technique or speaking ability could be obtained, would be to use as listeners a group who would remain constant in listening skill from test to test. A group of listeners, who had been trained until they had reached, or approached closely, the asymptote of their curve of listening skill, would meet this requirement. No such group has been available for our experiments, and subjects who could be thus trained and used as a criterion listening group over a period of time could not be obtained. However, rather careful consideration of this problem has brought us to the conclusion that it is not serious. After all, the important objective of training is to achieve the maximum over-all improvement in communication efficiency. It probably matters very little that some of that improvement results from incidental training to listen in noise. The question would seem to be more academic than practical. There is, however, a certain amount of evidence which points to the conclusion that most of the improvement found is on the speaking side. In some experiments, where different training methods and procedure have been tried out, definite and sizeable differences have been found between certain methods. These differences frequently have not been in the direction that one would have predicted on the hypothesis that all, or the bulk, of the improvement found was due to increased skill in listening. In fact, in some cases, the methods which have been found to be less effective were those in which the subjects had the most opportunity to listen to other subjects under conditions of noise, and hence to get practice which would improve their listening skill.

RESULTS FOLLOWING TRAINING

Results obtained from three typical training experiments will be presented and discussed. The basic data from these experiments are summarized in Tables I, II and III. The column headings Initial Test and Final Test are self-explanatory. N indicates the number of subjects composing the experimental training or control groups. The columns headed Gain present the raw improvement scores, obtained by subtracting the initial test mean from the final test mean. The data in the Adjusted Gain columns are these raw improvement scores adjusted for group differences in initial test score.

Table I. Results Obtained from Word Intelligibility Tests for Three Experimental Training Groups and a Control Group. Group Means for Initial Test, Final Test, Gain, and Adjusted Gain. Training Period - Two Hours

	N	Initial Test	Final Test	Gain	Adjusted Gain
Method A	41	41.7	52.2	10.5	11.0
Method B	39	40.8	50.8	10.0	10.1
Method C	35	42.7	51.8	9.1	9.9
Control	41	37.8	43.1	5.3	4.1

Table II. Results Obtained from Word Intelligibility Tests for Six Experimental Training Groups. Group Means for Initial Test, Final Test, Gain, and Adjusted Gain. Training Period - Three Hours

	N	Initial Test	Final Test	Gain	Adjusted Gain
Method A	28	50.1	59.0	8.9	9.0
Method B	26	55.4	61.5	8.1	10.5
Method C	27	46.0	59.4	13.4	10.5
Method D	31	52.4	61.5	9.1	10.8
Method E	30	45.0	63.8	18.8	15.1
Method F	27	53.9	67.4	13.5	16.3

Table III. Results Obtained from Word Intelligibility Tests for Four Experimental Training Groups and a Control Group. Group Means for Initial Test, Final Test, Gain, and Adjusted Gain. Training Period - Six Hours

	N	Initial Test	Final Test	Gain	Adjusted Gain
Method A	34	50.4	70.6	20.2	18.8
Method B	35	53.8	71.1	17.3	18.1
Method C	37	54.9	69.0	14.1	15.7
Method D	35	49.3	70.1	20.8	18.7
Control	38	54.1	58.9	4.8	5.8

This adjustment is necessitated by the fact that a consistent tendency exists for the amount of gain, made by an individual or a group, to be related to the initial test score of the individual or group. The nature of this relation is that of a negative correlation between initial score and gain. Other things equal, the relatively poor individuals or groups on the initial test make greater gain than do those who score relatively high on the initial test (see Table VI and Fig. 5). These differences in improvement are no doubt real, but when the problem is to evaluate the relative effectiveness of different training procedures, or the relative improvement of trained groups and control groups, the difficulty arises that the comparisons to be made are affected by differences in initial status of the particular groups involved. A different assignment of procedures to groups might produce different apparent advantage in unadjusted gain for the groups compared. One method of avoiding this difficulty is to insure that groups are relatively well matched at the beginning of the experiment. This can be done either by transferring subjects from one group to another, until the groups are fairly well matched, or by training all subjects up to a given level of proficiency before the start of the experiment. Either method insures relative equality of groups at the outset of the experiment. However, the first of these is administratively impossible, in the present instance, and the second is too time consuming. Fortunately, where the correlation between the control variable (Initial Score in this instance) and the criterion variable (Gain Score) is reasonably rectilinear,* this matching can be done statistically. Through the use of analysis of covariance, the regression of the initial test scores on gain can be used to adjust the gain scores to allow for initial differences between groups.** These adjusted gain scores are the figures found in the column headed Adjusted Gain.

The way in which this adjustment operates may be made clearer by an illustration. In Table II it will be noted that the group trained by Method C had a gain score, before adjustment, of 15.4 score points, while the group trained by Method D had an unadjusted gain score of 9.1 score points. These data, taken by themselves, would seem to indicate an advantage for Method C. It will be noted, however, that the Method C group had a relatively low initial test mean (46.0), while the initial test mean of the Method D group was substantially higher (52.4). Because of the negative correlation between initial score and gain, we know that this fact alone would give an advantage to the Method C group with respect to their unadjusted gain score, quite apart from any differences in effectiveness of the training methods that were used. Until the appropriate adjustment of these gain scores is made, therefore, it is not possible to come to any conclusion concerning the relative merits of the two methods. As will be seen from the table, the adjusted gain scores for these two groups show the difference between the two methods to have been negligible.

---

* Tests of linearity of regression between the initial scores and the gain scores have been applied for all experiments. In no case has there been any departure from linearity which approached significance.

** Cf. Lindquist, E. F., Statistical Analysis in Educational Research, Houghton-Mifflin, New York, 1940, Chap. VI.

The adjusted gain scores provide the best index by which to evaluate the improvement of the various experimental training and control groups. These adjusted gain scores are given graphically in Figures 2, 3, and 4.

Further reference to Tables I - III gives some indication of the relative amount of improvement in intelligibility that may be expected with training. It should be remembered that all of these groups had been given preliminary instruction concerning the use of the equipment before the initial test was administered. They had been instructed with respect to the proper way to hold the microphone and proper fitting of the headsets. Recently published results from another laboratory* have shown that substantial improvement in intelligibility may result, with untrained subjects, from very brief instruction in proper holding of the microphone. Improvement from this source has been purposely eliminated in the present experiments, by the preliminary instructions. Hence, the improvement shown by the data may be attributed almost entirely to the training received by these subjects in voice technique.

Although these experiments were not designed to evaluate the relative effectiveness of various lengths of training time, and there are too many differences between the experiments other than the time factor to permit accurate conclusions on this matter, some rough indications are shown. The best of the three methods for the groups trained for a two-hour period (Table I) yielded an adjusted gain of eleven score points. Among the groups trained for three hours, the best two methods produced adjusted gains of fifteen to sixteen score points. Among the groups trained for six hours, the best two groups show an adjusted gain of almost nineteen score points. It would seem, therefore, that there is an increase in improvement with increased training time. An experiment, now in progress, has been designed to give a more careful determination of this function. Results will be published shortly.

It has already been pointed out that the control groups also tended to show a certain amount of improvement, and possible reasons for that have been discussed. It can be seen, however, in Tables I and III and in Figures 2 and 4 that the trained groups, in all cases, showed substantially greater improvement than did the controls. A statistical analysis of these comparisons between the trained groups and their respective controls is presented in Tables IV and V. Table IV shows that as little as two hours of training produced improvement greater than that of the control group, by amounts which were statistically significant beyond the 1% level of confidence. The groups trained for six hours (Table V) compare even more favorably with their control group.

---

* Factors Related to the Intelligibility of Talkers in Noise, IC-60, Psycho-Acoustic Laboratory, Harvard University.

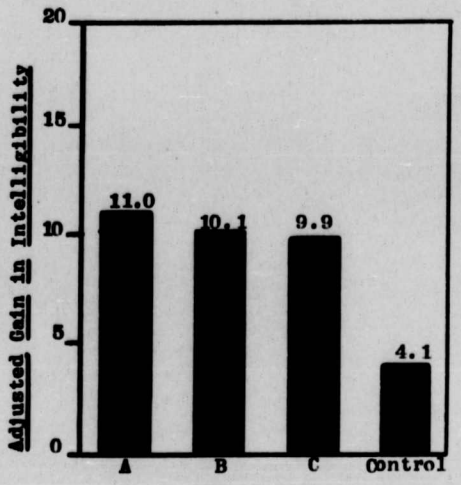

Figure 2. Comparison Between Adjusted Mean Gain of Three Groups Trained for a Period of Two Hours and a Control Group.

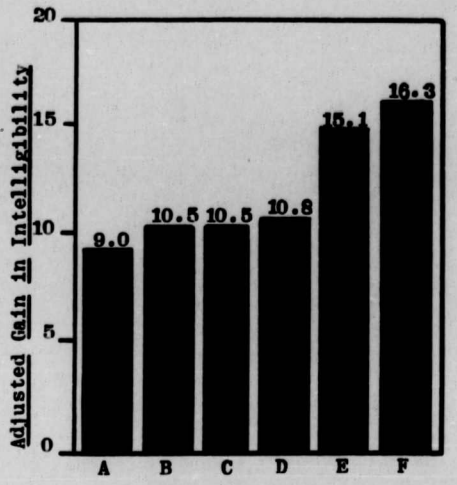

Figure 3. Comparison Between Adjusted Mean Gains of Six Groups Trained for a Period of Three Hours.

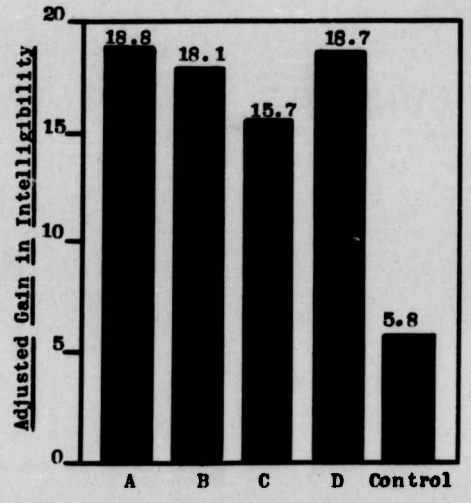

Figure 4. Comparisons Between Adjusted Mean Gains of Four Groups Trained for a Period of Six Hours and Control Group.

Table IV. Comparison of Improvement Found for Groups Trained for Two Hours with Control Group

Comparison	Difference in Adjusted Gain	Sigma of Difference	t
Method A v. Control	6.9	2.14	5.245
Method B v. Control	6.0	2.17	2.751
Method C v. Control	5.8	2.23	2.614

Value of t required for significance at the 1% level of confidence - 2.576.

Table V. Comparison of Improvement of Groups Trained for Six Hours with Control Group

Comparison	Difference in Adjusted Gain	Sigma of Difference	t
Method A v. Control	13.0	2.05	6.579
Method B v. Control	12.5	2.01	6.104
Method C v. Control	9.9	1.99	4.950
Method D v. Control	12.9	2.01	6.388

Value of t required for significance at the 1% level of confidence - 2.576.

The data of Table II and Figure 3 were included mainly to show that there are differences of appreciable magnitude among training procedures. Methods E and F yielded substantially greater improvement than did the other four. Statistical analysis shows these differences to be significant. It is not the intent of this report to discuss in detail the specific procedures involved. That material will be presented in other reports. The important point is that, for maximum results, careful consideration must be given to the particular training procedures to be employed.

Table VI and Figure 5 present data showing improvement by quartiles for groups trained for a period of six hours and for their control group. For this table all trained groups in Table III have been combined. The quartile division was made on the basis of the initial test scores.

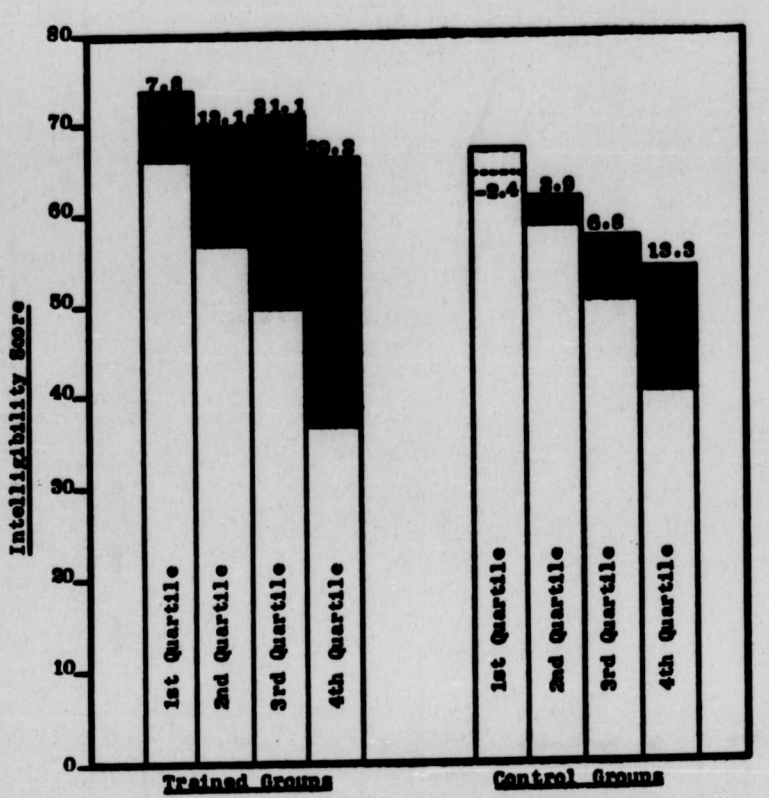

Figure 5. Improvement by Quartiles of Groups, Trained for Six Hours, Compared with Improvement by Quartiles of a Control Group. The white portions of the bars represent the initial test means. The white portions plus the cross-ruled portions represent the final test means, except for the 1st quartile of the control group. In this case there was a loss in score from the initial to the final test, and the portion of the bar below the dotted line represents the final test mean. The gains in intelligibility score, represented by the cross-ruled portions, are given in numerals at the tops of the bars.

Table VI. Improvement by Quartiles as a Result of Training. Groups Trained for Six Hours Compared to a Control Group

	TRAINED GROUPS			CONTROL GROUP			
Quartile*	Initial Test Mean	Final Test Mean	Gain	Initial Test Mean	Final Test Mean	Gain	Difference in Gain
1st	65.8	73.6	7.8	66.8	64.4	-2.4	10.2
2nd	56.9	70.0	13.1	58.9	61.8	2.9	10.2
3rd	49.5	70.6	21.1	50.8	57.6	6.8	14.3
4th	36.9	66.1	29.2	40.7	54.0	13.3	15.9

The considerable increase in improvement from the first to the fourth quartiles should be noted. Previously, attention was called to this relationship, in the discussion concerning adjustment of gain scores. This tendency for the initially poorer individuals to show the greater gain is not an unusual phenomenon in learning experiments; but it appears to be sufficiently marked, in this instance, to deserve special mention. The same tendency is shown by the control group. The first quartile of the control group actually showed a loss in intelligibility from the initial to the final test. However, the N for each quartile, in this group, is too small to permit any importance to be attached to this small change in score.

It will also be noted that, although rather wide differences exist among the four quartiles on the initial test score, most of these differences have disappeared on the final test. On the final test, the mean score of the fourth quartile is only 7.5 score points below the mean score of the first quartile, whereas on the initial test, there was a difference of 28.9 score points.

Probably one of the best comparisons, to indicate what training can accomplish, may be obtained from the data of Table VI by studying the performance of the lower two quartiles. It is reasonable to suppose that these initially poorer subjects, unless they receive adequate training, will be responsible for a major portion of poor communication later. It will be noted that the mean score for the lowest quartile on the final test, following training, is actually higher than the mean score of the upper quartile preceding training, and is equally as high as the mean score for the upper quartile of the control group, on either the initial or the final test.

* The division into quartiles was made on the distributions of initial test scores.

Stated somewhat differently, the average score of the individuals in the poorest twenty-five per cent of the subjects, prior to training, has increased, as a result of six hours of training, to the point where they are the equal, on the average, of the highest twenty-five per cent of untrained speakers. The mean score of the third quartile following training is even higher. These figures indicate that training can contribute considerably towards increasing communication efficiency.

DISCUSSION

A statement ought to be made concerning the characteristics of the subjects who have served in the above reported experiments. They were pilot cadets in the first four weeks of basic training at the Waco Army Air Field. They may be presumed to be a fair sample of pilot cadets throughout Army Air Forces. It is the unanimous opinion of the laboratory, including men who have been teachers of speech in colleges and universities for some years, that the ordinary speech of the average cadet was good, certainly above the average of the general population. Occasionally a subject would be classed as poor, who had bad articulation or poor pronunciation, or a bad voice, but such men were exceptions, and rather rare ones. Some of the men, indeed, had speech which would be classed as superior. More than a few had taken courses in speech in high school or college. Two or three had been radio announcers. Yet even these men with superior speech showed improvement in intelligibility for speaking in noise, following training, by amounts that are statistically significant. This is important in evaluating the need for this type of training. It means that good speech, as one ordinarily thinks of it, is not sufficient to insure maximum intelligibility over the interphone and radio under conditions of loud noise. Intelligible speaking under these conditions is a special problem, demanding special training.

It should be remembered that the training with which these experiments have been concerned has been training in voice technique only. All the results that have been reported have been based on training procedures whose objective has been to teach the men to speak and use their voices in a manner which would increase their intelligibility as much as possible. A well-rounded course in voice communication would, of course, include more than that. It would give the men practice in proper use of equipment and correct use of standard radiotelephone procedure. Many reports reaching this laboratory have indicated that there is considerable room for improvement in the latter of these, especially. Adequate training in this regard ought to provide an increase in the efficiency of voice communication which does not appear at all in the above reported results.

A final point concerns the practicability of fitting a training program in voice communication into the already crowded curricula of Army Air Forces training schools. It has been shown that significant improvement, compared to control subjects can be produced with as little as two hours

of training. The longest training period for any of the experiments, herein reported, was six hours. The addition of training in radiotelephone procedure and use of equipment might increase this time a slight amount. However, there is reason to believe that the training in voice technique can, to some extent, be effectively combined with training in standard procedure. In fact, most of the drill and practice materials used in the laboratory training experiments have been messages phrased in standard procedure language. It is the opinion of the laboratory that an effective training program, combining training in all of these skills, could be offered in six hours. Even shorter periods would be expected to produce substantial improvement, but maximal results could scarcely be expected in less time. It would seem, therefore, that not only has it been demonstrated that training in voice communication can be effective, but the desired results can be obtained without exceeding practical time limitations.

OVER

ATI- 35398

REVISION (None)

ORIG. AGENCY NO. R-8

PUBLISHING AGENCY NO. OSRD-3862

TITLE: Report on Training Studies in Voice Communication: I. Can Intelligibility of Voice Communication be Increased by Training in Voice Technique

AUTHOR(S): Curtis, James F., Brackett, I. P., Moser, H. M. and others

ORIGINATING AGENCY: The Psychological Corporation, New York, N. Y.

PUBLISHED BY: Office of Scientific Research and Development, NDRC, Div APP

DATE	DOC. CLASS.	COUNTRY	LANGUAGE	PAGES	ILLUSTRATIONS
July 44	[redacted]	U.S.	Eng.	20	tables, diagrs

ABSTRACT:

Data are given from three selected experiments to show the results obtained from training student pilot subjects in voice technique. In all of the training experiments, the criterion for evaluating results was a word intelligibility test. Results showed that training for a period as short as two hours produces gains in intelligibility score which are greater than that for a control group by amounts which are statistically significant at high levels of confidence. Further training up to six hours showed additional improvement. Significant differences exist between training procedures, indicating a need for careful working out of most advantageous procedures. A breakdown of the data into quartiles on the basis of the initial test scores shows that the subjects who scored lowest on the initial tests make the most marked improvement.

DISTRIBUTION: Copies of this report obtainable from Air Documents Division; Attn: MCIDXD

DIVISION: Education and Training (13) 3 4

SECTION: Flight Training (4)

SUBJECT HEADINGS: Communications, Voice - Intelligibility tests (23995.405)

ATI SHEET NO.: R-38-4-3

UNCLASSIFIED PER AUTHORITY OSRD LIST #9,
DATED 18-21 FEBRUARY 1946
BY John E. Moore, USCO